CLAIMS

We claim:

1.   A process for creating an ensemble filter for selecting documents, comprising:

identifying a set of documents for training;

identifying a first coherent set of documents from said training set of documents;

identifying a first profile corresponding to said first coherent set of documents;

identifying a second coherent set of documents and a remainder set of documents from said training set of documents using said first profile;

identifying at least one coherent set of documents from said remainder set of documents;

identifying at least one remainder profile corresponding to each of said identified coherent sets of documents from said remainder set of documents;

creating a first sub-filter using said first profile;

creating at least one remainder sub-filter using at least one of said remainder profiles; and

combining said first sub-filter with at least one remainder sub-filter to create an ensemble filter.

2.   A process, as in claim 1, further comprising:

clustering said training set of documents to identify said first coherent set of documents.

3.   A process, as in claim 1, further comprising:

clustering said training set of documents and selecting said largest cluster to

identify said first coherent set of documents.

4.   A process, as in claim 1, further comprising:

cascading said first sub-filter and at least one remainder sub-filter to create at least

part of said ensemble filter.

5.   A process, as in claim 1, further comprising:

mutiplexing said first sub-filter with at least one remainder sub-filter to create at

least part of said ensemble filter.

6.   A process, as in claim 2, further comprising:

cascading said first sub-filter and at least one remainder sub-filter to create at least

part of said ensemble filter.

7.   A process, as in claim 3, further comprising:

cascading said first sub-filter and at least one remainder sub-filter to create at least

part of said ensemble filter.

8.   A process, as in claim 2, further comprising:

mutiplexing said first sub-filter with at least one remainder sub-filter to create at least part of said ensemble filter.

9.   A process, as in claim 3, further comprising:

mutiplexing said first sub-filter with at least one remainder sub-filter to create at least part of said ensemble filter.

10.   A process for selecting documents from a stream of documents, comprising:

identifying a set of documents for training;

identifying a first coherent set of documents from said training set of documents;

identifying a first profile corresponding to said first coherent set of documents;

identifying a second coherent set of documents and a remainder set of documents from said training set of documents using said first profile;

identifying at least one coherent set of documents from said remainder set of documents;

identifying at least one remainder profile corresponding to each of said identified coherent sets of documents from said remainder set of documents;

creating a first sub-filter using said first profile;

creating at least one remainder sub-filter using at least one of said remainder profiles;

combining said first sub-filter with at least one remainder sub-filter to create an ensemble filter; and

passing said stream of documents through said ensemble filter.

11.   A process, as in claim 10, further comprising:

clustering said training set of documents to identify said first coherent set of

documents.

12.   A process, as in claim 10, further comprising:

clustering said training set of documents and selecting said largest cluster to

identify said first coherent set of documents.

13.   A process, as in claim 10, further comprising:

cascading said first sub-filter and at least one remainder sub-filter to create at least

part of said ensemble filter.

14.   A process, as in claim 10, further comprising:

mutiplexing said first sub-filter with at least one remainder sub-filter to create at

least part of said ensemble filter.

15.   A process, as in claim 11, further comprising:

cascading said first sub-filter and at least one remainder sub-filter to create at least

part of said ensemble filter.

16.   A process, as in claim 12, further comprising:

cascading said first sub-filter and at least one remainder sub-filter to create at least part of said ensemble filter.

17.   A process, as in claim 11, further comprising:

mutiplexing said first sub-filter with at least one remainder sub-filter to create at least part of said ensemble filter.

18.   A process, as in claim 12, further comprising:

mutiplexing said first sub-filter with at least one remainder sub-filter to create at least part of said ensemble filter.

19.   A process for selecting documents from a database of documents, comprising:

identifying a set of documents for training;

identifying a first coherent set of documents from said training set of documents;

identifying a first profile corresponding to said first coherent set of documents;

identifying a second coherent set of documents and a remainder set of documents from said training set of documents using said first profile;

identifying at least one coherent set of documents from said remainder set of documents;

identifying at least one remainder profile corresponding to each of said identified coherent sets of documents from said remainder set of documents;

creating a first sub-filter using said first profile;

creating at least one remainder sub-filter using at least one of said remainder profiles;

combining said first sub-filter with at least one remainder sub-filter to create an ensemble filter; and

applying said ensemble filter to said database to select documents.

20. A process, as in claim 19, further comprising:

clustering said training set of documents to identify said first coherent set of documents.

21. A process, as in claim 19, further comprising:

clustering said training set of documents and selecting said largest cluster to identify said first coherent set of documents.

22. A process, as in claim 19, further comprising:

cascading said first sub-filter and at least one remainder sub-filter to create at least part of said ensemble filter.

23. A process, as in claim 19, further comprising:

mutiplexing said first sub-filter with at least one remainder sub-filter to create at least part of said ensemble filter.

24. A process, as in claim 20, further comprising:

        cascading said first sub-filter and at least one remainder sub-filter to create at least part of said ensemble filter.

25. A process, as in claim 21, further comprising:

        cascading said first sub-filter and at least one remainder sub-filter to create at least part of said ensemble filter.

26. A process, as in claim 20, further comprising:

        mutiplexing said first sub-filter with at least one remainder sub-filter to create at least part of said ensemble filter.

27. A process, as in claim 21, further comprising:

        mutiplexing said first sub-filter with at least one remainder sub-filter to create at least part of said ensemble filter.